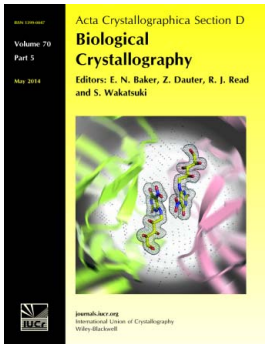
Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

A novel β-xylosidase structure from *Geobacillus thermoglucosidasius*: the first crystal structure of a glycoside hydrolase family GH52 enzyme reveals unpredicted similarity to other glycoside hydrolase folds

Giannina Espina, Kirstin Eley, Guillaume Pompidor, Thomas R. Schneider, Susan J. Crennell and Michael J. Danson

Acta Cryst. (2014). D**70**, 1366–1374



Acta Crystallographica Section D: Biological Crystallography welcomes the submission of papers covering any aspect of structural biology, with a particular emphasis on the structures of biological macromolecules and the methods used to determine them. Reports on new protein structures are particularly encouraged, as are structure–function papers that could include crystallographic binding studies, or structural analysis of mutants or other modified forms of a known protein structure. The key criterion is that such papers should present new insights into biology, chemistry or structure. Papers on crystallographic methods should be oriented towards biological crystallography, and may include new approaches to any aspect of structure determination or analysis. Papers on the crystallization of biological molecules will be accepted providing that these focus on new methods or other features that are of general importance or applicability.

Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

A novel β-xylosidase structure from *Geobacillus thermoglucosidasius*: the first crystal structure of a glycoside hydrolase family GH52 enzyme reveals unpredicted similarity to other glycoside hydrolase folds

Giannina Espina,[a] Kirstin Eley,[b] Guillaume Pompidor,[c] Thomas R. Schneider,[c] Susan J. Crennell[a]* and Michael J. Danson[a]

[a]Centre for Extremophile Research, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, England, [b]TMO Renewables Ltd, Alan Turing Road, Surrey Research Park, Guildford GU2 7YF, England, and [c]European Molecular Biology Laboratory (EMBL) Hamburg c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany

Correspondence e-mail: s.j.crennell@bath.ac.uk

Received 14 November 2013
Accepted 6 February 2014

PDB references:
β-xylosidase, 4c1o; complex with substrate, 4c1p

Geobacillus thermoglucosidasius is a thermophilic bacterium that is able to ferment both C6 and C5 sugars to produce ethanol. During growth on hemicellulose biomass, an intracellular β-xylosidase catalyses the hydrolysis of xylo-oligosaccharides to the monosaccharide xylose, which can then enter the pathways of central metabolism. The gene encoding a *G. thermoglucosidasius* β-xylosidase belonging to CAZy glycoside hydrolase family GH52 has been cloned and expressed in *Escherichia coli*. The recombinant enzyme has been characterized and a high-resolution (1.7 Å) crystal structure has been determined, resulting in the first reported structure of a GH52 family member. A lower resolution (2.6 Å) structure of the enzyme–substrate complex shows the positioning of the xylobiose substrate to be consistent with the proposed retaining mechanism of the family; additionally, the deep cleft of the active-site pocket, plus the proximity of the neighbouring subunit, afford an explanation for the lack of catalytic activity towards the polymer xylan. Whilst the fold of the *G. thermoglucosidasius* β-xylosidase is completely different from xylosidases in other CAZy families, the enzyme surprisingly shares structural similarities with other glycoside hydrolases, despite having no more than 13% sequence identity.

1. Introduction

β-Xylosidases (xylan 1,4-β-xylosidases; EC 3.2.1.37) catalyse the hydrolysis of (1,4)-β-D-xylo-oligosaccharides (*i.e.* xylobiose, xylotriose and xylotetrose) from their nonreducing termini into individual xylose units. β-Xylosidases are gaining increasing attention owing to their potential to reduce costs and environmental impact in several industrial and biotechnological processes. Their applications include the paper and pulp industry (*e.g.* bio-bleaching recycled paper and processing wood pulp; Suurnäkki *et al.*, 1997), the food industry (*e.g.* the production of xylitol, improving bread dough in baking, and winemaking), oligosaccharide and thioglycoside synthesis (Jordan & Wagschal, 2010) and second-generation biofuel production from renewable resources such as lignocellulose.

The efficient conversion of lignocellulosic biomass into liquid fuels is a major industrial goal, both to reduce dependency on fossil fuels and to reduce CO_2 emissions. To this end, the thermophilic ethanologen *Geobacillus thermoglucosidasius* has been metabolically engineered to divert carbon flux from a mixed-acid fermentation to one in which ethanol is the major product (Cripps *et al.*, 2009). However, the xylan component of lignocellulosic biomass remains

a microbially recalcitrant polymer. Xylan is a heteropolysaccharide containing different substituent groups in the backbone and side chains (Sunna & Antranikian, 1997), with the most common xylan substituents being arabinofuranose, glucuronic acid, methylglucoronic acid and acetyl groups. Therefore, a range of carbohydrases are necessary for the efficient utilization of typical biomass. However, once broken down into xylo-oligosaccharides through physical and enzymic treatment, intracellular β-xylosidases play a key role in converting these to the monosaccharide xylose, which can then be metabolized by organisms such as *G. thermoglucosidasius*.

To date, β-xylosidases have been classified according to their amino-acid sequence similarities in the Carbohydrate-Active enZymes database (CAZy; Cantarel *et al.*, 2009) within ten different glycoside hydrolase families: GH1, GH3, GH30, GH39, GH43, GH51, GH52, GH54, GH116 and GH120. High-resolution structures are available for representatives of all these families except for families GH52 and GH116. Czjzek *et al.* (2004) described a preliminary crystallographic analysis of a thermostable family 52 β-xylosidase from *G. stearothermophilus*, although to date no structure has been reported.

This paper reports the recombinant production, characterization and determination of both a high-resolution crystal structure and a lower resolution structure of the substrate complex of a *G. thermoglucosidasius* β-xylosidase that belongs to the CAZy glycoside hydrolase family GH52. This is the first reported structure of a member of family GH52, and the positioning of the substrate xylobiose in the active site is consistent with the retaining mechanism proposed for this family. Moreover, the two-domain protein was found to have high structural similarity to other glycosyl hydrolases with which it shares less than 13% sequence identity. The evolutionary relationships of these enzymes are discussed.

2. Materials and methods

2.1. Cloning of the β-xylosidase (GH52) gene

G. thermoglucosidasius strain TM242 (*ldh*−; *pdh*^{up}; *pfl*−; Cripps *et al.*, 2009) was grown aerobically at 60°C for 16 h in Tryptone Soya broth (Oxoid, Basingstoke, England) and its chromosomal DNA was isolated using standard techniques (Sambrook & Russell, 2001). The β-xylosidase gene (RTMO 01743) was PCR-amplified using Phusion HF DNA polymerase (Thermo Fisher Scientific, Loughborough, England) and the oligonucleotide primers 5′-CGGCTAGCATGCCAAAAAACATGTTTTTTAAC-3′ (forward) and 5′-CCGCTCGAGTTACTTTCCCTCTTCCAAC-3′ (reverse). Following amplification, the PCR product was purified, digested with *Nhe*I and *Xho*I restriction enzymes (Promega, Southampton, England) and then ligated into the expression vector pET-28a(+) (Novagen, Millipore, Watford, England) between the *Nhe*I/*Xho*I sites of the multiple cloning site to incorporate an N-terminal His tag. The sequence was confirmed to be correct.

2.2. Recombinant protein expression

Escherichia coli BL21(DE3) cells (Novagen) were transformed with the pET-28a(+)-β-xylosidase plasmid. Transformants were grown in TB medium [1.2%(*w*/*v*) tryptone, 2.4%(*w*/*v*) yeast extract, 72 m*M* K_2HPO_4, 17 m*M* KH_2PO_4, 0.4%(*v*/*v*) glycerol] containing kanamycin (50 µg ml^{-1}) at 37°C with shaking at 220 rev min^{-1} until an OD_{600} of 1 was reached, at which point protein expression was induced for 5 h by the addition of 0.3 m*M* IPTG (Melford, Ipswich, England). The cells were harvested by centrifugation at 6000*g* for 15 min at 4°C and resuspended in 10 ml 50 m*M* Tris–HCl buffer pH 7.4, 500 m*M* NaCl containing one EDTA-free protease-inhibitor tablet (Roche, Welwyn Garden City, England). The cells were lysed by five 15 s bursts of sonication using a 150 W Ultrasonic Disintegrator (MSE Scientific Instruments, Crawley, England). A soluble cell extract was obtained by centrifugation at 14 000*g* for 20 min at 4°C and was filtered through a 0.22 µm filter (Millipore Ltd) prior to purification.

2.2.1. Selenomethionine-labelled protein. *E. coli* BL21 cells harbouring the pET-28a(+)-β-xylosidase plasmid were grown overnight in 5 ml LB medium containing kanamycin (50 µg ml^{-1}) at 37°C with shaking at 220 rev min^{-1}. The cells were collected by centrifugation, washed twice with sterile distilled water and then inoculated into 500 ml sterile SelenoMethionine Medium Complete (Molecular Dimensions Ltd, Newmarket, England) in 2 l baffled flasks. The cells were grown at 37°C with shaking at 220 rev min^{-1} until an OD_{600} of 1 was reached. The medium was then supplemented with lysine, phenylalanine and threonine at 100 mg l^{-1} and isoleucine, leucine and valine at 50 mg l^{-1} to block methionine biosynthesis in *E. coli* (Doublié, 2007). 2 ml SelenoMethionine Solution (250×; 10 mg ml^{-1}; Molecular Dimensions) were then added 15 min prior to induction with 0.3 m*M* IPTG for 5 h. The cells were harvested, resuspended and lysed as before, adding 1 m*M* tris-(2-carboxyethyl)phosphine (Calbiochem, Millipore) to the resuspension buffer.

2.3. Protein purification

β-Xylosidase was purified by nickel-affinity chromatography and gel filtration at 25°C on an ÄKTAexplorer FPLC system (GE Healthcare, Chalfont St Giles, England). The soluble cell lysate in 50 m*M* Tris–HCl buffer pH 7.4 containing 500 m*M* NaCl was loaded onto a pre-equilibrated HisTrap 5 ml HP column (GE Healthcare) at a flow rate of 1 ml min^{-1}. The column was washed with the loading buffer followed by the same buffer containing 60 m*M* imidazole; finally, the β-xylosidase was eluted with Tris–HCl buffer pH 8.0, 300 m*M* NaCl, 200 m*M* imidazole.

Fractions containing β-xylosidase activity were pooled and concentrated using an Amicon Ultra 10K centrifugal filter device (Millipore) and then further purified by gel filtration on a Superdex 200 10/300 GL column (GE Healthcare) at a flow rate of 0.2 ml min^{-1} in 50 m*M* Tris–HCl pH 8 containing 150 m*M* NaCl. Fractions were analysed by SDS–PAGE and those containing pure protein were pooled and concentrated

to 20 mg ml^{-1} using an Amicon Ultra 10K centrifugal filter device.

Protein concentrations for the pure enzyme were determined from $A_{280\ nm}$ readings using an absorption coefficient of 144 785 M^{-1} cm^{-1} as determined by *ProtParam* (Gasteiger *et al.*, 2005).

2.4. Enzyme assays

β-Xylosidase activity was routinely assayed spectrophotometrically using the chromogenic substrate *p*-nitrophenyl-β-D-xylopyranoside (Glycosynth, Warrington, England). All assays were conducted in a volume of 1 ml at 65°C using preheated McIlvaine's buffer (0.16 *M* Na_2HPO_4, 0.018 *M* citric acid pH 7.0) and the appearance of *p*-nitrophenol was monitored at 410 nm; the molar absorption coefficient determined under these conditions was 8500 M^{-1} cm^{-1}. Catalytic activity with *p*-nitrophenyl-α-L-arabinopyranoside (Glycosynth Ltd), *p*-nitrophenyl-α-L-arabinofuranoside (Carbosynth, Compton, Berkshire, England) and *p*-nitrophenyl-β-D-glucopyranoside (Sigma–Aldrich, Gillingham, England) was determined under the same conditions. All assays were initiated by the addition of enzyme, and one unit of β-xylosidase activity is defined as the release of 1 µmol of *p*-nitrophenol per minute.

Catalytic activity with the natural substrates xylobiose and xylotriose was assayed by determination of the xylose produced using NAD^+-xylose dehydrogenase as a coupling enzyme from the Megazyme D-Xylose Assay Kit (Megazyme, Bray, Ireland). Spectrophotometric assays (1 ml) were carried out at 45°C in Megazyme buffer, and the production of NADH was followed continuously at 340 nm (the molar absorption coefficient of NADH is 6220 M^{-1} cm^{-1}). One unit of β-xylosidase activity is defined as the production of 1 µmol of NADH per minute.

Assay data were analysed using the Enzyme Kinetics module in *SigmaPlot* 12 (Systat Software, Hounslow, England).

2.5. Mass spectrometry

Selenomethionine (SeMet) incorporation into the protein was verified by mass spectrometry. Native and SeMet-derivatized protein at 5 mg ml^{-1} concentration were analysed using a Nano LC-CHIP-MS system consisting of a 1200 Nano HPLC-Chip microfluidic device (Agilent Technologies, Santa Clara, California, USA) coupled to an Agilent 6520 Accurate-Mass Quadrupole-Time-of-Flight Liquid Chromatography-Mass Spectrometry (Q-TOF LCMS).

A 1 µl sample was loaded from the autosampler onto the Agilent Chip enrichment column [5 µm ZORBAX 300SB-C3 (300 Å), 40 nl] at a loading flow of 4 µl min^{-1}, followed by passage to the analytical column [5 µm ZORBAX 300SB-C3 (300 Å), 75 µm × 43 mm] at 300 nl min^{-1}. The composition of mobile phase *A* was 0.1%(*v*/*v*) formic acid, and the liquid phase *B* consisted of 90%(*v*/*v*) acetonitrile and 0.1%(*v*/*v*) formic acid. Elution was obtained with increasing concentration of buffer *B*: 3% for the first 8 min, 50% for 2 min and finally 100% for 2 min. The percentage of buffer *B* was then decreased to 3% and maintained at this level to allow column re-equilibration.

Samples were analyzed by ESI in positive-ion mode. Mass-spectrometric data were acquired in the *m*/*z* range 100–3000 with an acquisition rate of 1.35 spectra per second, averaging 10 000 transients. The source parameters were adjusted as follows: drying gas temperature 300°C, drying gas flow rate 5 l min^{-1}, nebulizer pressure 310 kPa and fragmentor voltage 150 V. Data acquisition and processing were performed using the *Agilent MassHunter Workstation* acquisition software (B.02.01 Build 2116).

2.6. Structural studies

2.6.1. Protein crystallization. Crystallization was performed in 24-well plates by the hanging-drop vapour-diffusion method at 18°C. Native crystals, as well as crystals of protein derivatized with selenomethionine, were obtained from β-xylosidase (10 mg ml^{-1} in 50 m*M* Tris–HCl buffer pH 8, 150 m*M* NaCl) mixed in equal volumes with well solution consisting of 0.1 *M* MES buffer pH 6.0, 4 *M* ammonium sulfate, 25% PEG 3350. Large crystals of approximately 0.1 × 0.05 × 0.05 mm in size appeared after three weeks at 18°C. Glycerol [10%(*v*/*v*) diluted in well solution] was used as cryoprotectant when the crystals were flash-cooled prior to data collection.

Co-crystallization of β-xylosidase with its product xylose (100 m*M*) was performed using the same conditions, although sodium formate (1.6 *M* in well solution containing 100 m*M* xylose) was used as a cryoprotectant.

2.6.2. X-ray data collection and processing. Diffraction data were collected on EMBL beamline P13 at PETRA III (DESY, Hamburg, Germany). The beamline was equipped with a PILATUS 6M-F detector (DECTRIS, Baden, Switzerland) and an MD2 microdiffractometer (MAATEL, Moirans, France). Both SeMet-derivatized and native data sets were collected at λ = 0.9763 Å (12.700 keV), slightly above the Se *K* edge (12.658 keV), in shutterless data-collection mode. For the SeMet-derivative crystals a wedge of 540° was collected from a single crystal with an exposure time of 250 ms per 0.2° oscillation (25% beam transmission) at a crystal-to-detector distance of 343 mm.

The native data set was measured over a total wedge of 140° with angular steps of 0.2°, an exposure time of 100 ms with a beam transmission of 100% and a crystal-to-detector distance of 319 mm. The data were processed using *XDS* (Kabsch, 2010) and were scaled with *SCALA* (Evans, 2006). Diffraction data for the native enzyme co-crystallized with xylose were collected in-house at 100 K using a Rigaku MicroMax-007 HF X-ray generator with a Saturn 944+ charge-coupled device detector at a wavelength corresponding to the Cu $K\alpha$ edge (1.5418 Å). The data were processed using the *HKL*-2000 package (Otwinowski & Minor, 1997). Data-collection and processing statistics for both data sets are presented in Table 1.

2.6.3. Structure solution and refinement. The structure was solved *via* SAD phasing using the SeMet anomalous diffraction data with beta-test versions of *SHELXC*/*D*/*E* (Sheldrick, 2008) *via* a beta-test version of *HKL2MAP* (Pape &

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Glycerol	Formate + xylose
Data-collection statistics		
Space group	$C222_1$	$C222_1$
Unit-cell parameters (Å, °)	a = 73.56, b = 105.10, c = 195.60, $\alpha = \beta = \gamma$ = 90	a = 73.369, b = 105.335, c = 194.834, $\alpha = \beta = \gamma$ = 90
Resolution (Å)	1.7 (1.73–1.70)	2.64 (2.69–2.64)
Total No. of reflections	429091	508871
No. of unique reflections	83431	22164
Completeness (%)	99.9 (100.0)	97.3 (78.0)
Multiplicity	5.1 (5.3)	5.0 (2.1)
$\langle I/\sigma(I)\rangle$	18.6 (2.7)	7.62 (1.73)
$R_{\rm merge}$	0.045 (0.530)	0.138 (0.481)
Overall average B factor from Wilson plot (Å^2)	19.6	14.9
Structure refinement		
Resolution range (Å)	97.78–1.70	27.31–2.634
No. of reflections		
Working set	83366	22136
Test set	4165	1169
Final $R_{\rm cryst}$	0.1404	0.1823
Final $R_{\rm free}$	0.1730	0.2380
No. of non-H atoms		
Protein	5862	5674
Ligands/ions	167/21	26/1
Water	637	306
R.m.s. deviations		
Bonds (Å)	0.010	0.003
Angles (°)	1.261	0.805
B factors (Å^2)		
Protein	23.55	13.75
Ligand/ions	47.56/69.10	20.69/10.63
Water	35.59	12.5
Ramachandran plot analysis, residues in		
Most favoured regions (%)	97.5	96.01
Disallowed regions	0	1
MolProbity score	1.12 [99th percentile; N = 9248, 1.70 ± 0.25 Å]	1.58 [99th percentile; N = 6042, 2.63 ± 0.25 Å]
Clashscore	2.64 [99th percentile; N = 819, 1.70 ± 0.25 Å]	2.63 [100th percentile; N = 226, 2.63 ± 0.25 Å]

Schneider, 2004). For solution of the selenium substructure, 20 sites were requested and the data were truncated to 3.0 Å resolution; for density modification a solvent content of 46% was specified, but default values were used for all other parameters. The electron-density map resulting from *SHELXE* after refinement of the selenium sites and ten macro-cycles consisting of 20 cycles of density modification followed by auto-tracing was automatically interpreted by *ARP*/*wARP* (Langer *et al.*, 2008), building 697 residues.

Manual model building was performed with *Coot* (Emsley *et al.*, 2010), refinement was performed using *REFMAC5* (Murshudov *et al.*, 2011) and *PHENIX* (Adams *et al.*, 2010), and final model evaluation was performed using *MolProbity* (Chen *et al.*, 2010).

The product-bound structure was solved by molecular replacement using the unbound structure as a model, and was rebuilt and refined as above. Both structures have been deposited in the PDB as entries 4c1o and 4c1p, respectively. Surface interface areas were determined using the *PISA* server (*Protein Interfaces, Surfaces and Assemblies*) at the European Bioinformatics Institute (Krissinel & Henrick, 2007). PDB searches and structural comparisons were performed using the *DALI* server (Holm & Rosenström, 2010) and *PDBeFold* (Krissinel & Henrick, 2004).

3. Results and discussion

3.1. Characterization of the recombinant β-xylosidase

The recombinant β-xylosidase from *G. thermoglucosidasius* TM242 was expressed in a soluble, catalytically active form in *E. coli* and was purified to homogeneity by nickel-affinity and gel-filtration chromatography. Mass-spectrometric analysis gave a molecular mass of 81 980, suggesting that the N-terminal methionine had been removed (the expected molecular mass minus methionine is 81 975). Size-exclusion chromatography with standard proteins of known molecular masses gave an molecular mass of 168 000, indicating that the enzyme is a dimer.

Using *p*-nitrophenyl-sugar derivatives as substrates, it is clear that the enzyme has a strong preference for β-D-xylopyranoside substrates, although there was also significant activity with the α-L-arabinopyranoside derivative, albeit with a very high $K_{\rm m}$ (Table 2). The enzyme has excellent activity with the natural substrates xylobiose and xylotriose; however, these assays had to be carried out at 45°C owing to the mesophilic nature of the coupling enzyme, xylose dehydrogenase. From assays of the temperature dependence of enzyme activity using pNP-β-D-xylopyranoside, the activity of the enzyme with xylobiose (24.5 µmol pNP min^{-1} mg^{-1}) would equate to approximately 100 µmol pNP min^{-1} mg^{-1} at 65°C and a $k_{\rm cat}$ of 135 min^{-1}. Similar adjustments to the values with xylotriose give 68 µmol pNP min^{-1} mg^{-1} at 65°C and a $k_{\rm cat}$ of 92 min^{-1}. No activity was found with xylan, thus confirming the identity of the enzyme as a β-xylosidase. The temperature optimum for the enzyme was determined to be 65°C, at which temperature the half-life was 50 min as measured by its time-dependent irreversible thermal inactivation. These data are those expected for an enzyme from *G. thermoglucosidasius*, which grows optimally at 60°C.

The selenomethionine-derivatized enzyme was also catalytically active (its specific activity using pNP-β-D-xylopyranoside as the substrate was 38 U mg^{-1}) and the determined molecular mass of 82 864 by mass spectrometry indicated that all of the methionine residues had been replaced by selenomethionine (expected molecular mass of 82 868).

3.2. β-xylosidase structure

The structure determined by SAD has two domains: an N-terminal domain folding into a β-sandwich and a C-terminal α/α_6 barrel. Analysis using the EBI *PISA* server (Krissinel & Henrick, 2007) indicated that the β-xylosidase exists as a stable dimer within the crystal, the symmetry-related molecule being rotated so the N-terminal domain of one molecule interacts with the C-terminal domain of the other, burying

Table 2
Catalytic activity and kinetic parameters of the recombinant β-xylosidase.

Errors on all values are <10%. Assays with the artificial substrates were carried out at 65°C and those with the natural substrates at 45°C.

(*a*) Artificial substrates.

	V_{max} (µmol pNP min^{-1} mg^{-1})	k_{cat} (s^{-1})	K_m (m*M*)	k_{cat}/K_m (mM^{-1} s^{-1})
pNP-β-D-xylopyranoside	40.5	55.4	0.34	163
pNP-β-D-glucopyranoside	1.3	1.8	12.8	0.1
pNP-α-L-arabinopyranoside	7.6	10.4	27.4	0.4
pNP-α-L-arabinofuranoside	0.3	0.4	0.8	0.5

(*b*) Natural substrates.

	V_{max} (µmol NADH min^{-1} mg^{-1})	k_{cat} (s^{-1})	K_m (m*M*)	k_{cat}/K_m (mM^{-1} s^{-1})
Xylobiose	24.5	33.5	2.4	13.9
Xylotriose	16.9	23.1	3.5	6.6

18 370 Å^2 of surface, with ten hydrogen bonds and four salt bridges across the interface (Fig. 1).

As the cryoprotectant, glycerol, can take up a conformation mimicking half of a sugar ring, the position of the active site was suggested by a glycerol molecule bound tightly in a cleft in the α/α_6 domain and on the interface between the two monomers (the average *B* factor for this glycerol was less than half of the mean *B* factor for the 17 glycerols bound in the structure; 22.4 Å^2 compared with 48.3 Å^2). A second glycerol molecule was bound outside the cleft but close to the first tightly bound glycerol (average *B* factor of 32.2 Å^2) with hydrogen bonds being formed across the dimer interface, indicating a possible second sugar-binding site. Consistent with this is our finding that glycerol is a competitive inhibitor of the β-xylosidase activity (data not shown).

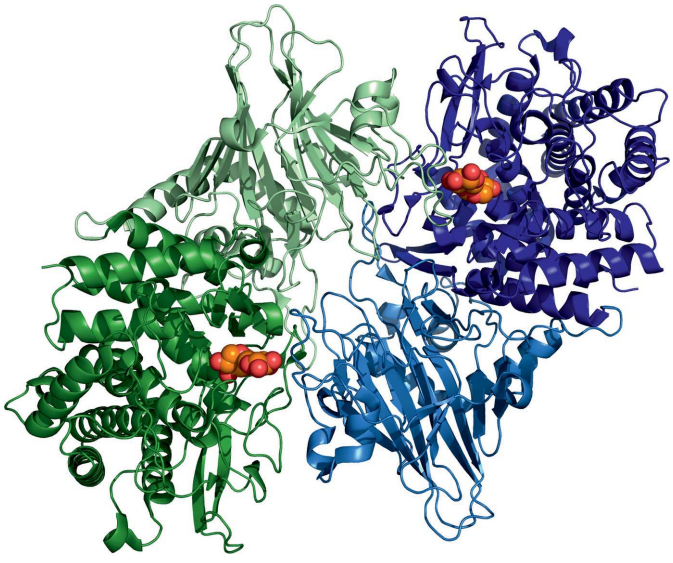


Figure 1
A cartoon representation of the β-xylosidase dimer. Each N-terminal domain is shown in lighter colours and each C-terminal domain is shown in darker colours. Xylobiose is shown in space-filling representation in the active sites.

Since the glycerol cryoprotectant is required at a concentration that would compete with substrate or product binding to the active site, formate was used as the cryoprotectant in the xylose co-crystal data collection. This structure confirmed the cleft as the predominant xylose-binding site, and sufficient electron density was seen to fit two xylose molecules into the proposed active site; however, the continuity of the density is indicative of these two xylose units having been combined into a single molecule of xylobiose (Fig. 2). It should be noted that xylose was in considerable excess over its K_m and therefore it is entirely possible that the enzyme has catalysed the reverse, glycosynthetic, reaction.

Xylose did not bind anywhere else in the structure. The overlapping positioning of glycerol (cryoprotectant) and the

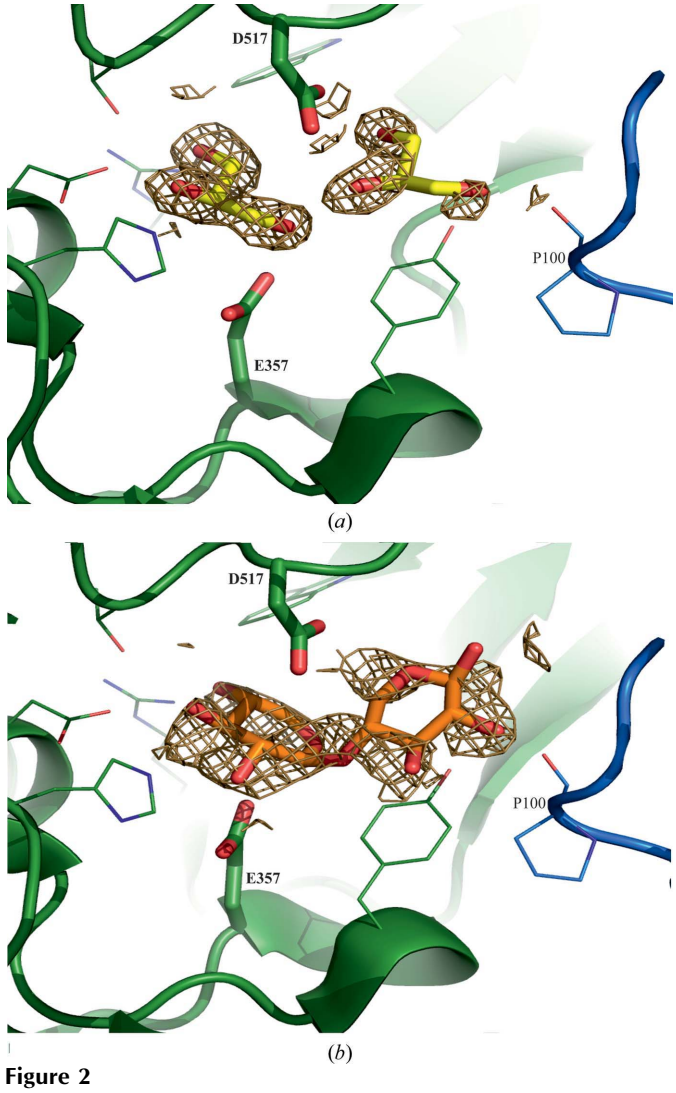


Figure 2
The β-xylosidase active site with the residues forming the active-site cleft shown in stick form and the catalytic residues shown in thicker lines. The proline on the right-hand side of the active site comes from the partner subunit of the dimer. (*a*) Glycerol bound in the active site of the native structure, with the simulated-annealing OMIT electron density within 1.5 Å of the glycerol molecules contoured at 1.2σ. (*b*) Xylobiose bound in the active site of the native structure overlaid with the simulated-annealing OMIT electron density within 1.5 Å of the xylobiose molecule contoured at 1.2σ.

Table 3
Interactions shorter than 3.25 Å made by xylobiose in the xylosidase active site, and comparison with families 39 (data from Table 1 of Czjzek *et al.*, 2005) and 120 (Huang *et al.*, 2012).

Xylobiose	GH52 BXP	Distance (Å)	GH39 XYP/ANX	Distance (Å)	GH120 BXP	Distance (Å)
XylA stacking	Tyr360		Tyr283		His352	
O1A					Water	
O2A	Pro100 O†	2.46				
O3A					Glu405 $O^{\varepsilon2}$	3.17
C4	Asp517 $O^{\delta2}$	3.13			Glu405 $O^{\varepsilon2}$	2.59
O4A			Glu160 $O^{\varepsilon2}$	2.84		
(O1 in ANX)			Glu160 $O^{\varepsilon1}$	2.75		
C1B	Glu357 $O^{\varepsilon2}$	2.99	Glu278 $O^{\varepsilon1}$	3.22‡	Asp382 $O^{\delta2}$	3.10
O2B	Glu357 $O^{\varepsilon2}$	3.11	Glu160 $O^{\varepsilon2}$	2.90	Asp382 $O^{\delta1}$	2.76
	Glu357 $O^{\varepsilon1}$	3.14	Glu278 $O^{\varepsilon2}$	2.40	Arg450 $N^{\eta1}$	2.82
	His418 $N^{\varepsilon2}$	2.99	Asn159 $N^{\delta2}$	2.75		
			Asn159 $O^{\delta1}$	2.88		
O3B	Asp367 $O^{\delta2}$	2.89	Glu324 $O^{\varepsilon2}$	2.62	Trp383 $N^{\varepsilon1}$	2.93
			His54 $N^{\varepsilon2}$	2.38	Gln289 $O^{\varepsilon1}$	2.62
					Water	2.95
O4B	Thr515 $O^{\gamma1}$	3.00	Glu324 $O^{\varepsilon1}$	2.53	Glu353 $O^{\varepsilon2}$	2.57
	Arg715 $N^{\eta1}$	3.00	Trp316 $N^{\varepsilon1}$	3.18	Glu353 $O^{\varepsilon1}$	3.09
	Trp654 $C^{\zeta3}$	3.10			Lys358 N^{ζ}	3.17
					His360 $N^{\varepsilon2}$	2.83
					Asp382 $O^{\delta2}$	3.18
O5B	Glu357 $O^{\varepsilon2}$	3.12	Tyr230 O^{η}	3.25	Asp382 $O^{\delta2}$	3.00
					Lys358 N^{ζ}	3.13
					His352 O	3.16
XylB stacking					Trp113	

xylobiose in the active site confirms that glycerol is mimicking the natural substrate (Fig. 2).

There are two distinct xylose-binding subsites, one of which is buried and accommodates the −1 nonreducing end of xylose, whilst the second is more open, with its main interaction with the +1 reducing end of xylose being a nondirectional stacking interaction with Tyr360. The −1 nonreducing end of xylose is held in an enclosed pocket by several hydrogen bonds (Table 3) and is distorted into a higher energy $^4H_3$ half-chair conformation, in marked contrast to the low-energy chair conformation taken up by the +1 xylose.

A single hydrogen bond is made across the active-site cleft between O2A (of the +1 xylose) and a main-chain atom from the other molecule in the β-xylosidase dimer, showing a restricted access to the active site that prevents larger xylan polymers from entry (Fig. 3).

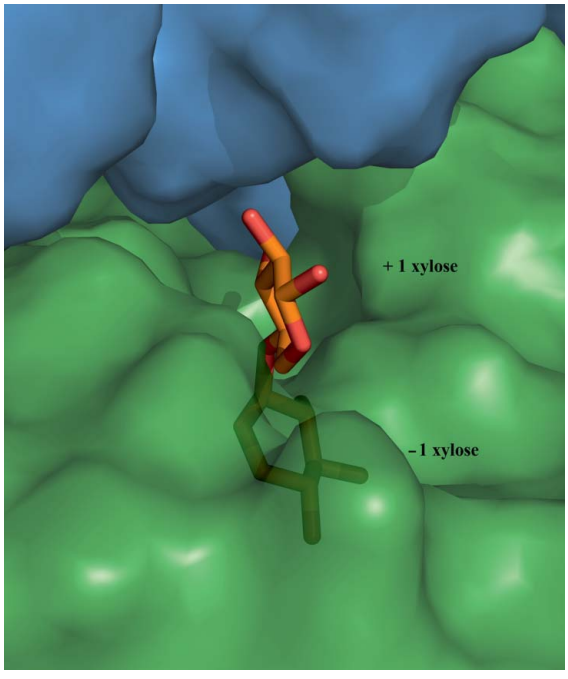


Figure 3
A surface representation, with 20% transparency, of the β-xylosidase active-site cleft with the xylobiose substrate shown in stick form. The two subunits of the dimeric enzyme are shown in green and blue.

3.3. The catalytic site

Glycosides fall into two mechanistic classes: one hydrolysing the glycosidic bond with net inversion of the anomeric configuration and one doing so with net retention (McCarter & Withers, 1994). Both classes of enzyme utilize a pair of acidic residues, which are on average 9.0–9.5 Å apart in inverting enzymes but only 4.8–5.3 Å apart in those with a retaining mechanism. Previous studies have shown that β-xylosidases belonging to the family 52 glycoside hydrolases possess a retaining mechanism of catalysis in which hydrolysis of xylo-oligosaccharides to xylose occurs with net retention of configuration (Bravman *et al.*, 2001). This is achieved *via* a two-step double-displacement mechanism involving a covalent glycosyl-enzyme intermediate, with each step passing through an oxocarbenium ion-like transition state (Koshland, 1953).

Within the active-site pocket of the *G. thermoglucosidasius* β-xylosidase, the two carboxylic residues that are positioned across the glycosidic O between the xylose units, Glu357 and Asp517, are 6.5 Å apart (Fig. 2), indicative of the enzyme

displaying a retaining mechanism of catalysis. Furthermore, in a sequence alignment with the XynB2 β-xylosidase from *G. stearothermophilus* T-6 (GenBank code ABI49956.1), these two residues align perfectly with Glu335 and Asp495 of this enzyme. Through site-directed mutagenesis of the *G. stearothermophilus* enzyme, followed by detailed kinetic analyses including the pH-dependence of the catalytic constants, Bravman *et al.* (2003) have demonstrated that Glu335 functions as a nucleophile in the glycosyl-cleavage reaction, with Asp495 assisting as an acid. Asp495 then acts as a base to catalyse the hydrolysis of the glycosyl-glutamyl intermediate. These, and the other amino acids shown to interact with xylobiose, are conserved in all GH52 sequences in CAZy, and a sequence alignment of representative genera is given in Supplementary Fig. 1[1]. The one exception is XylA from *G. stearothermophilus* 236, but this sequence is 60 amino acids shorter than any other members of the family, making alignments difficult.

3.4. Comparison with structures of enzymes with the same xylosidase function

β-Xylosidases, as identified by EC number, are found in nine glycoside hydrolase families in CAZy (Supplementary Table S1), the majority of which are of known fold, although most of the structures deposited are of proteins with other glycosidase activities (Supplementary Table S1). The available β-xylosidase structures are for members of families 39, 43 and 120, of which families 39 and 120 have mechanisms that retain product configuration, as does *G. thermoglucosidasius* β-xylosidase. The folds of these enzymes are completely different from that of the GH52 family described here, being a $(\beta/\alpha)_8$ barrel and β-sandwich in GH39 and a β-helix and a β-sandwich in family GH120 (Supplementary Table S1). However, comparison of their active sites suggests common features for protein active sites with β-xylosidase activity (Fig. 4).

Comparison of *G. thermoglucosidasius* β-xylosidase with the GH120 structure (PDB entry 3vsu; Huang *et al.*, 2012), which also has xylobiose bound, reveals that both have a deep pocket binding the −1 xylose with many hydrogen bonds, while the +1 xylose is bound much less strongly, with few hydrogen bonds and stacking interactions with an aromatic residue. In both cases the +1 site is in close proximity to another monomer of the oligomer, although interactions are only by a water-mediated hydrogen-bond network in GH120. The closer proximity of the other monomer in *G. thermoglucosidasius* β-xylosidase (Fig. 4) causes the active-site cleft to be significantly narrower than in the GH120 structure. In both cases, the −1 sugar takes up a skew conformation, but the +1 xylose is in the relaxed chair conformation in *G. thermoglucosidasius* β-xylosidase, while that in GH120 is in the higher energy boat form.

The GH39 (PDB entry 2bfg; Czjzek *et al.*, 2005) active site shares more similarities with the *G. thermoglucosidasius*

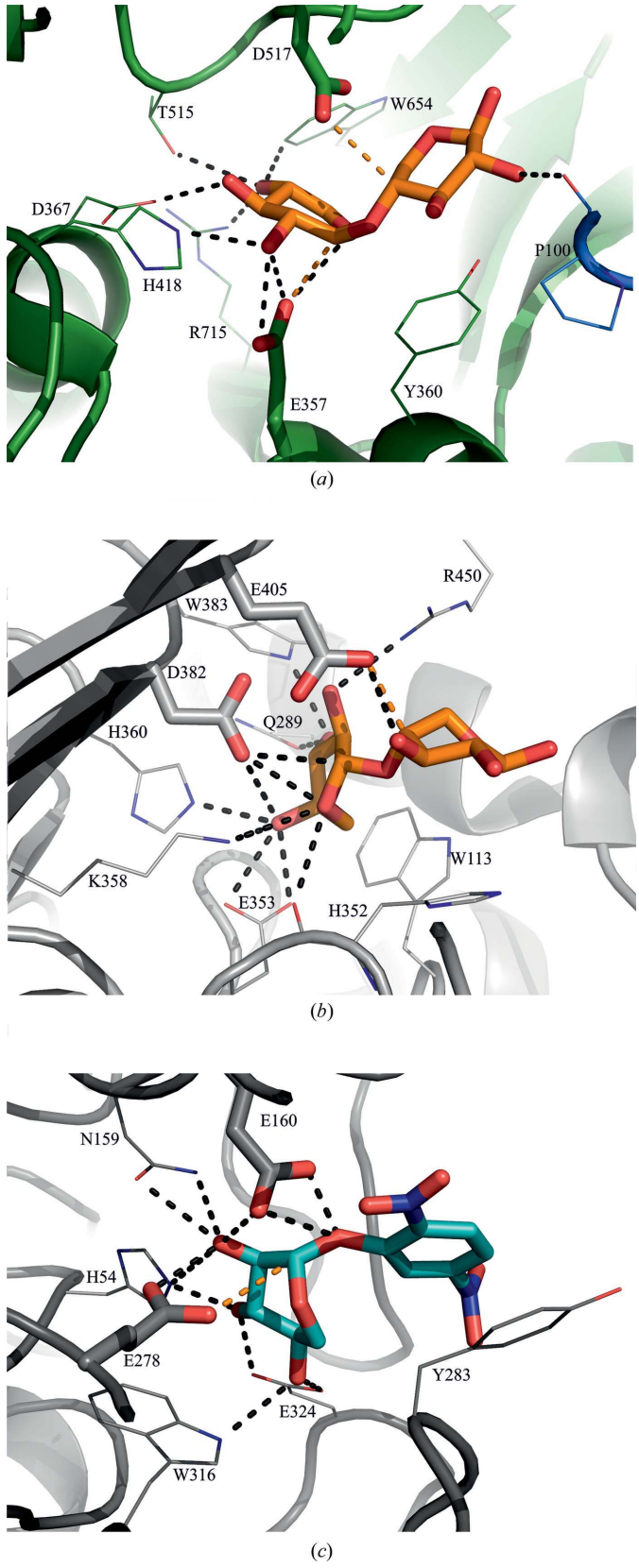


Figure 4
Interactions within the active sites of β-xylosidases. (*a*) *G. thermoglucosidasius*, (*b*) GH120 (PDB entry 3vsu) and (*c*) GH39 (PDB entry 2bfg). The substrate is shown in thick bonds and the interacting residues as sticks; hydrogen bonds are shown as black dashed lines and carbon–oxygen interactions of less than 3.2 Å are shown in orange.

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: WD5231).

β-xylosidase: both have a deep cleft to accommodate the substrate, restricting access for larger substrates, although this is within one monomer in GH39. If aligned by bound substrate, the nucleophiles (Glu278 and Glu357, respectively) also align, while the general bases do not. The buried -1 xylose is distorted into a similar conformation in both active sites, caused by a mixture of strong hydrogen bonds and hydrophobic interactions and a stacking interaction in GH39. The conformation of the +1 xylose cannot be determined as pNP occupies this site in 2bfg, but it is clear that both form stacking interactions with Tyr.

As in the *G. thermoglucosidasius* β-xylosidase (Fig. 3), solvent-accessibility analysis of the substrate bound in all three structures reveals that the -1 xylose is completely buried (data not shown), while the +1 xylose has variable levels of exposure to solvent. This lack of exposure of the -1 xylose to bulk solvent would render the enzyme incapable of cleaving bulk xylan, thus explaining the lack of xylanase activity.

3.5. Comparison with similar structures

Solving the *G. thermoglucosidasius* β-xylosidase structure by molecular replacement was not successful as no structures shared more than 20% sequence identity. However, once the structure had been solved, a search of the PDB using either the Protein Structure Comparison Service *PDBeFold* (Krissinel & Henrick, 2004) or the *DALI* server (Holm & Rosenström, 2010) showed that the *G. thermoglucosidasius* β-xylosidase shares striking structural similarities with other proteins, despite sharing no more than 13% sequence identity (Fig. 5).

As all of the *PDBeFold* results were also identified by *DALI*, only the *DALI* results are given in Supplementary Table S2. The structures identified as similar are usually glycoside hydrolases, from families GH15, GH63, GH65, GH78, GH92, GH94 and GH95. Of these, just GH15 and GH65 have a clan assigned, both belonging to the clan GH-L, which denotes an $(\alpha/\alpha)_6$ fold; this clan assignment would match the C-terminal domain of the *G. thermoglucosidasius* β-xylosidase. The classic position for an active site in an α/α_6 fold, in the groove at the centre of the barrel above the N-terminal ends of the inner helices, is also conserved in the β-xylosidase. In those structures containing both similar domains this location is conserved (Fig. 5). However, the dimeric form of the *G. thermoglucosidasius* β-xylosidase is not strictly conserved in the majority of these other families. Restriction of access to the C-terminal active site by the N-terminus of another monomer may be an

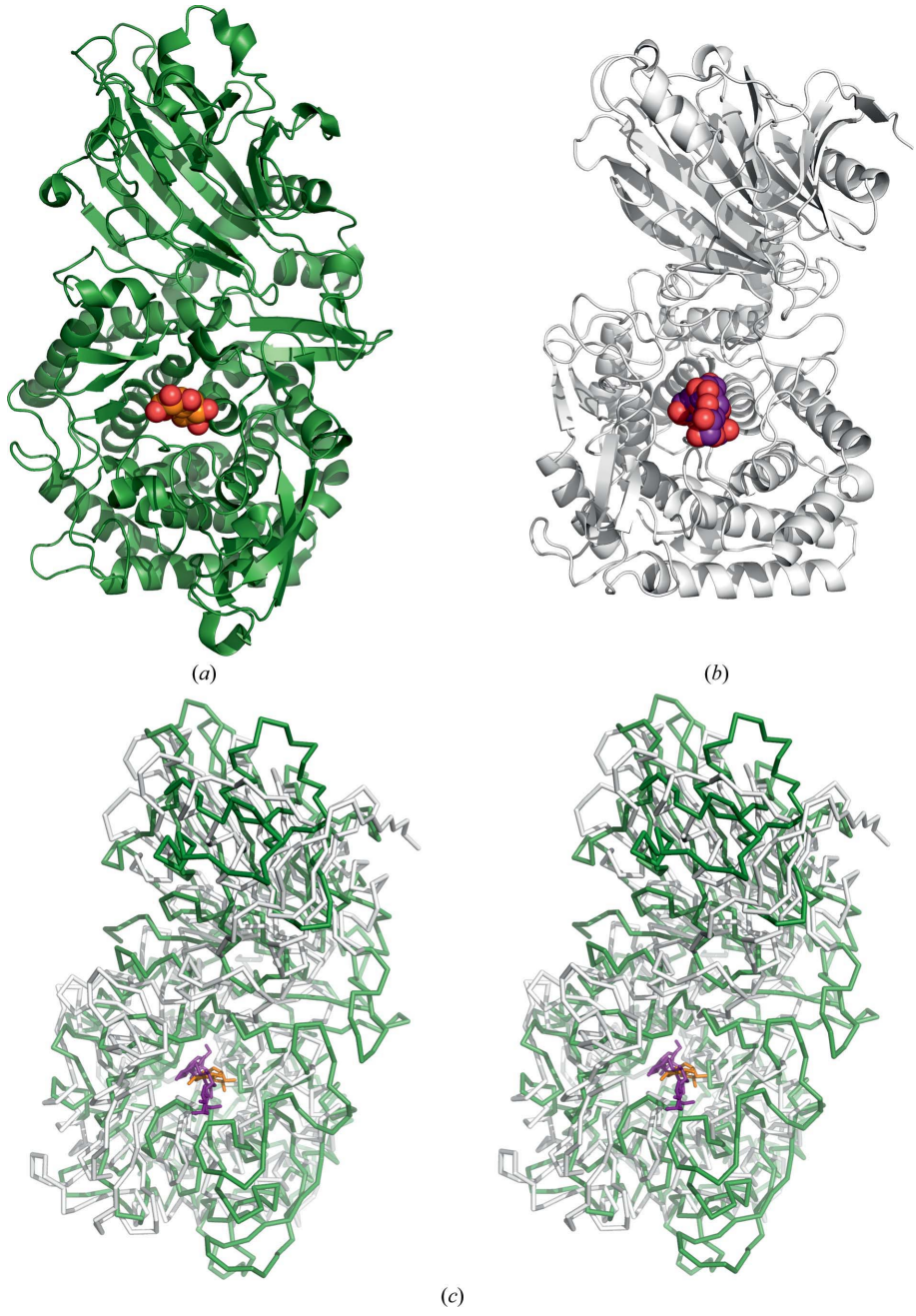


Figure 5
Structure alignment of *G. thermoglucosidasius* β-xylosidase (green) with a representative GH15 structure (white) chosen to be of similar size. (*a*) A cartoon representation of *G. thermoglucosidasius* β-xylosidase with xylobiose in space-filling spheres. (*b*) A cartoon representation of GH15 glucoamylase (PDB entry 1lf9) with acarbose (inhibitor) bound. (*c*). A stereoview of the C^{α} traces of *G. thermoglucosidasius* β-xylosidase and the aligned GH15 structure, with the xylobiose in β-xylosidase shown as orange sticks and the acarbose inhibitor in the glucoamylase shown in purple.

adaptation to restrict access to smaller xylo-oligosaccharides rather than bulk polymer xylan.

This identification suggests that an ancestral protein contained this combination of structural domains and has subsequently evolved to hydrolyse a wide range of glycosides to the extent that sequence identity is no longer discernible.

4. Related literature

The following references are cited in the Supporting Information: Canteral *et al.* (2009) and Gouet *et al.* (1999).

We thank Dr Rob Meijers (EMBL, Hamburg) and Dr Jean van den Elsen (University of Bath) for their expert advice, and Mr Mervyn Lewis (University of Bath) for assistance with the mass-spectrometric analysis. The research leading to the crystallization results received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement No. 227764 (P-CUBE; http://www.p-cube.eu/). We also indebted to the National Commission for Scientific and Technological Research (CONICYT) of Chile for a studentship to GE.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Bravman, T., Belakhov, V., Solomon, D., Shoham, G., Henrissat, B., Baasov, T. & Shoham, Y. (2003). *J. Biol. Chem.* **278**, 26742–26749.
Bravman, T., Zolotnitsky, G., Shulami, S., Belakhov, V., Solomon, D., Baasov, T., Shoham, G. & Shoham, Y. (2001). *FEBS Lett.* **495**, 39–43.
Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Cripps, R. E., Eley, K., Leak, D. J., Rudd, B., Taylor, M., Todd, M., Boakes, S., Martin, S. & Atkinson, T. (2009). *Metab. Eng.* **11**, 398–408.
Czjzek, M., Ben David, A., Bravman, T., Shoham, G., Henrissat, B. & Shoham, Y. (2005). *J. Mol. Biol.* **353**, 838–846.
Czjzek, M., Bravman, T., Henrissat, B. & Shoham, Y. (2004). *Acta Cryst.* D**60**, 1461–1463.
Doublié, S. (2007). *Methods Mol. Biol.* **363**, 91–108.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Huang, C.-H., Sun, Y., Ko, T.-P., Chen, C.-C., Zheng, Y., Chan, H.-C., Pang, X., Wiegel, J., Shao, W. & Guo, R.-T. (2012). *Biochem. J.* **448**, 401–407.
Jordan, D. B. & Wagschal, K. (2010). *Appl. Microbiol. Biotechnol.* **86**, 1647–1658.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Koshland, D. E. Jr (1953). *Biol. Rev.* **28**, 416–436.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
McCarter, J. D. & Withers, G. S. (1994). *Curr. Opin. Struct. Biol.* **4**, 885–892.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd ed., Vol. 2. Cold Spring Harbor Laboratory Press.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sunna, A. & Antranikian, G. (1997). *Crit. Rev. Biotechnol.* **17**, 39–67.
Suurnäkki, A., Tenkanen, M., Buchert, J. & Viikari, L. (1997). *Adv. Biochem. Eng. Biotechnol.* **57**, 261–287.